

Overview

This README file describes how to replicate the results of the paper “Normalizations and misspecification in skill formation models” by Joachim Freyberger.

Data availability statement

All raw data for this paper comes from Attanasio, Meghir, and Nix (2020). The data is publicly available on the Oxford University Press website:

<https://academic.oup.com/restud/article/87/6/2511/5839752#supplementary-data>

The data can be downloaded directly without access to the journal article using this [link](#).

The specific data used in my paper is contained in the file cohort1indianewr2.csv.

The application first uses cohort1indianewr2.csv to construct simulatedvariables.csv, which is used in the remaining analysis. This latter file is included in the replication package. Therefore, all figures can be created even without the replication package of Attanasio, Meghir, and Nix (2020) starting with step 3 below.

Software versions

R and MATLAB are needed to replicate the results.

I used RStudio with R version 4.2.2. The code requires the packages corpcor (I used version 1.6.10), gdata (I used 3.0.0), ks (I used 1.14.2), Matrix (I used 1.6-5).

I used MATLAB R2022b with the optimization and the statistics toolboxes.

Folder structure

The following folders and subfolders should exist in the replication package:

- ➡ Application
 - ➡ Figures
 - ➡ Get Data AMN
 - ➡ Code
 - ➡ Data
 - ➡ Output
- ➡ Figure Example
- ➡ Monte Carlo Simulations

Step-by-step instructions

1. Download the replication package of Attanasio, Meghir, and Nix (2020) using the link above.
 - (a) Copy the files `estim_meas_model5rev.R`, `Estimate_FactorModel.R`, `MeasurementSpec_wcalbookrevisions_healthgrowth.R`, and `Procedures.R` from their folder “Datafortheweb\Code” and place them in the folder “Application\Get Data AMN\Code” of this replication package.
 - (b) Copy the file `cohort1indianewr2.csv` from their folder “Datafortheweb\Data” and place it in the folder “Application\Get Data AMN\Data” of this replication package.
2. Run `Application_Main.R` in the folder “Application\Get Data AMN\Code”.
 - (a) The working directly in line 7 has to be adapted.
 - (b) The code creates the file `simulatedvariables.csv` in the folder “Application\Get Data” based on the code of Attanasio, Meghir, and Nix (2020).
 - (c) Since these draws might be slightly different when using a different version of R, the csv file with the draws my analysis is based on is contained in the replication package.
3. Run `Estimation_AMN_data_all_scales.m` in the folder “Application” in MATLAB (and to be run with the folder “Application” as the current directory).
 - (a) The code creates Figures 5, 6, 7, A.1, A.2, and A.3 and outputs them into the subfolder “Figures”.
 - (b) Lines 410-415 of the file `Estimation_AMN_data_original.m`, which is run as part of the code in `Estimation_AMN_data_all_scales.m`, contains the different interpretations of the effect of investment on skills in Section 5 of the paper.
4. Run `Simulations_Main.m` in the folder “Monte Carlo Simulations” in MATLAB (and to be run with the folder “Monte Carlo Simulations” as the current directory).
 - (a) The code creates Figure 2, 3, 4.
5. Run `plot_figure_1.m` in the folder “Figure Example” in MATLAB (and to be run with the folder “Figure Example” as the current directory).
 - (a) The code creates Figure 1.

Approximate run times

Using a desktop computer with an Intel Core i9 processor (2.8 GHz, 10 cores, 64 GB memory),

- step 2 takes around 15 minutes to run,
- step 3 takes around 20 minutes to run,
- step 4 takes around 40 hours to run, and
- step 5 takes less than a minute to run.

Acknowledgment

The research was supported by the European Research Council (ERC-2020-STG-949319).

Reference

Attanasio, Meghir, and Nix (2020), “Human Capital Development and Parental Investment in India“, *The Review of Economic Studies*, 2020, 87(6), 2511-2541.